Surface Features of a *Mononegavirales* Matrix Protein Indicate Sites of Membrane Interaction

Victoria A. Money[1], Helen K. McPhee[2♦], Jackie A. Mosely[2♦], John M. Sanderson[2*] and Robert P. Yeo[3*]

Addresses:

[1]The Maurice Wilkins Centre for Molecular Biodiscovery and the School of Biological Sciences, University of Auckland, Thomas Building, 3a Symonds Street, Auckland Central 1010, New Zealand;[2]Department of Chemistry and [3]Centre for Bioactive Chemistry, University Science Laboratories, South Road, Durham, DH1 3LE, UK

♦*These authors contributed equally to this work*

* Joint corresponding authors

Email and phone numbers

RPY: +44 (0)191 3340691, r.p.yeo@durham.ac.uk

JMS: +44 (0)191 3342107, j.m.sanderson@durham.ac.uk

Author Contributions; J.M.S & R.P.Y designed research; V.A.M, H.K.M & J.A.M performed the research; V.A.M, J.M.S. & R.P.Y analyzed data and wrote the paper

Coordinates and observed structure factors for the RSV M protein have been deposited with the Protein Data Bank (accession codes **2vqp** for the coordinates and **r2vqpsf** for the structure factors).

Abstract

The Matrix (M) protein of Respiratory Syncytial Virus, the prototype viral member of the *Pneumovirinae* (family *Paramyxoviridae*, order *Mononegavirales*), has been crystallized and the structure determined to a resolution of 1.6 Å. The structure comprises two compact β-rich domains connected by a relatively unstructured linker region. Due to the high degree of side chain order in the structure, an extensive contiguous area of positive surface charge covering approximately 600 $Å^2$ can be resolved. This unusually large patch of positive surface potential spans both domains and the linker and provides a mechanism for driving the interaction of the protein with a negatively-charged membrane surface or other virion components such as the nucleocapsid. This is complemented by regions of high hydrophobicity and a striking planar arrangement of tyrosine residues encircling the C-terminal domain. Comparison of the RSV M sequence with other members of the *Pneumovirinae* shows that regions of divergence correspond to surface exposed loops in the M structure, with the majority of viral species-specific differences occurring in the N-terminal domain.

\body

Introduction

Respiratory Syncytial Virus (RSV) is the prototype member of the *Pneumovirinae*, a subfamily of the *Paramyxoviridae* (order *Mononegavirales*). Morphologically the extracellular virion consists of a lipid bilayer envelope within which are embedded three glycoproteins, two of which (F and G) are important in cell attachment and viral entry into target cells; the third, the SH protein, contributes to pathology in the host (1). Internally, virions contain helical nucleocapsids that consist of N protein tightly bound to the negative-sense non-segmented genomic RNA. The nucleocapsid in turn is associated with components of the viral RNA-dependent RNA polymerase (L, P, M2-1 and M2-2 proteins) forming the holo-nucleocapsid (2-4). Between the holo-nucleocapsid and the outer envelope is a layer of matrix (M) protein which is associated peripherally with the membrane (5). The other family members of the *Mononegavirales* (*Rhabdoviridae*, *Filoviridae* and *Bornaviridae*) all subscribe to this basic arrangement of the virion, although the overall morphology can vary between the families; for example *Paramyxoviridae* virions are pleiomorphic whilst the *Rhabdoviridae* have a regular bullet shape structure and the *Filoviridae* have a more filamentous shape.

Extracellular RSV virions form by a budding process that occurs at the plasma membrane within specialized lipid domains (5, 6) and M appears to drive the final assembly process, which is the incorporation of the holo-nucleocapsid and initiation of the budding process (7, 8). Prior to budding there is a co-ordinated assembly of viral components and it is evident that the glycoproteins and M proteins are important determinants of the location on the plasma membrane at which the virus buds (9). It is also possible that the interaction between M proteins and the glycoproteins, possibly mediated with the cytoplasmic tails, are important in the budding process. Genome silencing to prevent transcription and replication by the viral polymerase prior to incorporation of the holo-nucleocapsid in the nascent virion is a related function of the M protein. M proteins are also implicated in host cell transcriptional cut-off (10) possibly via a direct interaction with RNA (11).

A number of matrix proteins are known to bind membranes or lipid vesicles *in vitro*, most likely though a combination of hydrophobic and electrostatic interactions (12-14). Expression of certain matrix proteins in eukaryotic cells in the absence of other viral proteins can induce formation of Virus-Like Particles (VLPs). The efficiency of VLP generation can be increased if the matrix protein is co-expressed with a viral glycoprotein (15-18). Matrix proteins share a tendency to oligomerize, a feature likely to be important in the self-assembly and budding processes (19). In tissue culture RSV induces formation of long slender projections from the surface of the cell known as viral filaments. It was found that removal of the lipid membrane from viral filaments left an M containing sheath (5); work has shown that the formation of such projections can be stopped by actin polymerization inhibitors, suggesting a role for actin, although immuno-staining for actin within these structures has not been able to demonstrate its presence (RPY unpublished observations). It is possible then, that M oligomerization and self-assembly is the driving force behind the formation of viral filaments.

In this paper we report the structure of the full-length RSV M protein solved at a resolution of 1.6 Å and discuss the implications of this structure for the function of the protein.

Results and Discussion

Structure of the RSV matrix protein (M)

The RSV M protein was purified using nickel affinity chromatography . During the cloning process a methionine to arginine change occurred and we refer to the resultant form of M as M^{254R}. The crystal structure of M^{254R} was solved using MIRAS techniques to a resolution of 1.6 Å, representing the first example of an intact matrix protein from the *Mononegavirale*s. Despite evidence for higher order oligomers (see supplementary information Figure S1) such as dimers, tetramers and hexamers, in solution the crystallized form is monomeric. Crystallographic data are presented in supplementary Table S1. The overall fold consists of two clear domains connected by a 13-residue linker region. The N-terminal domain comprises residues 1 to 126 whilst the C-terminal domain consists of residues 140 to 255 (Figure 1 a, b). Only residues 99 and 100 could not be assigned a clear location within the electron density. The N-terminal

domain consists of a twisted β-sandwich comprised of two β-sheets, one of three and one of four strands, positioned almost perpendicular to each other. The overall topology of this domain is of a curved horse-shoe like arrangement with β-sheet 1 forming the concave, inner face flanked by loop regions and β-sheet 2 forming the convex, outer face. The C-terminal domain consists of a flattened β-barrel comprising two three-stranded anti-parallel β-sheets. The regions linking the sheets between strands 2 and 3 and strands 5 and 6 are largely helical in nature (Figure 1 b). There is no evidence for any complexed metal ions or for a potential zinc finger motif. The 254R substitution in the protein lies at the very end of the C-terminal domain in an area largely devoid of secondary structure and is therefore unlikely to have a significant effect on the protein structure or function.

The linker region is largely lacking in secondary structure features with the exception of a short helical region. The presence of this linker is consistent with structures obtained for fragments of the Ebola virus (EBOV) VP40 and Vesicular Stomatitis Virus (VSV) M proteins, and from capsid proteins of retroviral origin that fulfil similar *in vivo* roles (20-24) and its unstructured nature suggests that the N- and C-terminal domains may be able to occupy different orientations, relative to each other, than that observed in the crystal. The two domains are only loosely associated, with the major interactions between them being hydrophobic in nature, supported by a small number of water-mediated hydrogen bonds. We have observed that, in common with EBOV VP40, proteolysis of RSV M^{254R} occurs in solution and results in the dissociation of the N- and C-terminal domains, the weak interdomain interactions being insufficient to hold them together. We have mapped the cleavage site of M^{254R} by limited mass spectrometry to the linker region between amino acids Thr136 and Leu137 (data not shown). These weak interdomain interactions further suggest that the protein may exist in alternative quaternary structures in solution and that the interdomain packing observed in the crystal may be metastable, driven mainly by sequestering of the hydrophobic residues, which form the major part of the interface, away from the bulk solvent. Flexibility of this type has been speculated as being important for the matrix protein of the influenza virus as well as for EBOV VP40 and VSV

M and is thought to be necessary to accommodate the different functions of the matrix protein throughout the viral life cycle.

Structure comparison

The tertiary structure of M^{254R} is globally similar to that of the Ebola VP40 protein (24), showing the same overall fold with a Z score of 6.6 and a rmsd of 3.7 Å. Searches of the PDB for structurally homologous proteins to M^{254R} using the programs DALI, VAST and SSM (25-27) yield no further significantly similar structures. Performing the same searches using the N- and C-terminal domains independently demonstrates that, in addition to being closely related to each other, the N-terminal domain shows slight similarities to parts of two DNA topoisomerases from *E. coli*. These similarities are restricted to the shape of the β-sheets in domain 2 of the DNA binding proteins and are unlikely to be relevant to the function of the N-terminal domain.

Structural information on matrix proteins is scarce partly due to the intrinsic difficulty of working with these hydrophobic proteins that are prone to self aggregation. Only two matrix proteins from the *Mononegavirales* have been subjected to high resolution structure determination. For both published structures, a proteolytically resistant core was crystallized as opposed to the full length structure. Separate overlay of the N- and C-terminal domains of M^{254R} with those of Ebola VP40, (Figure 2), allows the difference in angle between the domains in the two proteins to be accounted for, and shows the similarities in the protein core of each domain. Despite the clear close relation between the folds of the two proteins, the topological arrangement of the two differs. The N-terminal domain of RSV M^{254R} contains a mixed 4-stranded and one anti-parallel three-stranded β-sheet as opposed to the two three-stranded sheets found in VP40. A similar comparison of M^{254R} and VSV M could not be performed, emphasising the structural, but not functional, diversity of the two matrix proteins.

Electrostatic surface of the RSV Matrix protein

In order to fulfil its structural role, RSV M must be able to form protein-protein and protein-lipid interactions. Consequently, one would expect to observe surface areas with significant hydrophobic patches, as well as positively charged regions that would favor protein-

membrane association. Examination of the surface of M reveals an extensive positively charged area, of ca. 600 Å^2, extending across both N- and C-terminal domains encompassing, and including, a significant contribution from the linker (Figure 1 c). As the binding of M to cell membranes is thought to be mediated largely by electrostatic contacts (12, 28) this region provides a mechanism by which the protein is able to associate with negatively charged host membranes. Comparison of the electrostatic surface of M^{254} with VP40 is hampered by the fact that in the latter the linker region is not modelled. It does appear, however, that a significant positively charged patch is a feature present on all matrix proteins for which there is structural information available, including retroviral homologues. Looking at the rest of M^{254R}, there is a distinctive negatively charged lobe on the N-terminal domain and a positively charged pocket on the C-terminal domain, which are potential sites for directing interactions with binding partners. Work on the isolated N-terminal domain of Ebola matrix protein, VP40 (29), indicates that this domain has the capacity to oligomerize in the presence of nucleic acids, pointing to a more prominent role for this domain in RNP association, with the C-terminal domain interacting predominantly with the membrane. The precise role of each domain in the full length proteins (RSV M and EBOV VP40) has yet to be determined, although the extent of the positively-charged surface of RSV M indicates that functional surfaces have the potential to extend across the domain boundary and the linker to some degree.

Comparison of structures in solution and in crystal

In order to determine if the crystal structure was also that of the protein in solution, we performed a CD analysis of the protein. Data collected included wavelengths down to 190 nm, which significantly increases confidence in the interpretation of CD derived structural information for comparison with crystallographic data (29). Evaluation of CD (Figure 3 and Table 1, see supplementary information for more details of CD spectrum deconvolution) and X-crystallographic data indicate that the secondary structure content of M^{254R} in solution is lower than that observed in the crystalline state. This implies that assembly of the protein is associated with significant unfavorable entropic contribution from protein folding, in common

with the binding of disordered proteins to their counterparts (30-33). Such proteins achieve a favorable binding free energy through large enthalpic contributions from intramolecular electrostatic interactions to offset negative entropic contributions from protein folding. In order for this to be achieved, the number of contacts between the protein and its binding partners are numerous. In the case of M, analogy with disordered proteins would suggest that the contact area for protein-protein and protein-membrane interactions is similarly large. This is consistent with the observed distribution of aromatic and positively charged residues in the protein, particularly with regard to the large patch of positive electrostatic potential described above. Intrinsic disorder has been proposed as a mechanism by which viral proteins are able to form multiple binding interactions with different partners, thereby expanding protein functionality without a concomitant increase in the size of the genome. RSV M fits this description having multiple interactions; such as with itself, the nucleocapsid, the viral glycoproteins, such as the F protein, via their cytoplasmic tails and with the host cell and viral membranes. All of these are essential for the assembly and budding of a virion particle and could contribute to stabilizing the structure.

Comparison with other Pneumovirinae Matrix proteins

Analysis of sequence alignments of a number of pneumovirus and metapneumovirus M proteins with M^{254R} reveals that the majority of significant amino acid sequence diversity (Q-scores of 20 and above) can be mapped to external loop and edge regions of the β-sheets and to the linker region (Figure 4 a, b, see supplementary Figure S5 for alignments). An alignment of the more closely related bovine and ovine RSV M proteins with human RSV M proteins demonstrates that the major variations in amino acid sequence are mostly found in regions at each end of the horseshoe structure, with the remainder occurring in the linker and the ends of the helix that lies on the outer surface of the N-terminal domain. Inclusion of the more distantly related metapneumoviral M proteins in the alignment, such as those from human and avian metapneumoviruses, produces a similar pattern for the C-terminal domain; however, the differences are much more pronounced in the N-terminal domain, with significant sequence

diversity occurring in the linker and in the regions adjacent to the linker. We hypothesize that these differences correlate with surface residues that mediate species-specific interactions, such as M with its cognitive nucleocapsid or viral glycoproteins. This is consistent with the involvement of loop regions at the end of each domain in protein-protein and/or protein-RNA interactions. Given that membrane interactions are unlikely to be sequence-specific and that protein activity is not transferred between species outside experimental conditions, this is consistent with the involvement of loop regions in protein-protein and/or protein-RNA interactions, particularly for the N-terminal domain, where sequence diversity is greatest

A model for membrane binding

The large positively-charged area on the surface of RSV M, which spans both domains, is consistent with the role of this protein in membrane association. We hypothesize that it is likely that this patch will be the driving force for association with the negatively-charged lung membrane (34, 35). This would be consistent with biochemical observations of other M protein where electrostatic charge is the major component of the interaction between protein and membrane, at least *in vitro*. Association of the positively charged patch to the membrane, particularly at N- and C-terminal domain orientations other than those observed in the crystal structure, would leave a significant hydrophobic area on the protein exposed that may drive interactions with other viral components, or become buried at the protein-membrane interface.

Considering the C-terminal domain alone, while a significant proportion of the surface residues are hydrophobic (36) it is pertinent that this domain has a number of surface exposed arginine and lysine residues that are able to contribute to a favorable interaction with negatively-charged membranes. Membrane binding by the C-terminal domain, driven largely by electrostatic interactions, with a contribution from hydrophobic residues, would leave the N-terminus free to perform more species-specific functions, such as protein-protein interactions needed for virion assembly; this interpretation would be consistent with observations on EBOV VP40 (30). We also note that the C-terminal domain has a striking arrangement of tyrosine residues (Y163, Y197, Y215, Y229, Y236 and Y237) forming a planar distribution that encircles

the majority of the C-terminal domain (Figure 5 a, b) with a degree of overlap with the positively charged patch. Tyrosine, amongst other aromatic amino acids, has been demonstrated to occur with high probability in the interfacial region of membrane proteins (37-39) and in model systems has been shown to associate strongly with phospholipids (40). The functionality of the tyrosine residues has yet to be investigated but the motif is conserved within all RSV M proteins. The metapneumoviral M proteins do not have this motif, although in these proteins functionally equivalent residues replace the tyrosines. Mutagenesis studies based on the structure presented within this paper will further our understanding of the contribution of particular residues or features to the biological functionality of this protein.

Conclusion

The structure of the matrix protein of RSV has been solved by X-ray crystallography to a resolution of 1.6 Å. This is the first example of a full-length structure of a matrix protein from the viral order *Mononegavirales*. The high resolution and high degree of crystallographic order observed in the structure allow us to throw light on the mode of membrane binding and the mechanism by which this protein performs its varied and critical roles. Central to this is a significant area (~600 $Å^2$) of positive electrostatic potential that forms an extended surface for interaction with the membrane, which carries a complementary negative potential. As this area spans both domains and the linker, the geometry of surface contacts made by the protein will depend on the relative positioning of the domains, should those contacts involve contributions from both domains.

Experimental Procedures

Methods Summary

Protein expression purification and crystallization.

A histidine tagged version of M^{254R} was expressed in *E. coli* strain BL21 (CodonPlus). Cells were lyzed by sonication and M^{254R} purified by nickel-affinity chromatography. After removal of insoluble material the protein was subjected to crystallization trials. Successful crystallization

was obtained in 70% Tacsimate, pH 7.0, later the conditions were optimized to 55-65% Tacsimate, pH 7.0.

X-ray data collection

Data from native, seleno-methionine derivatized and mercury soaked crystals were collected at the ESRF (native crystal) and SSRL (derivatized forms) facilities. The crystal statistics are presented in the supplementary material accompanying this article. The three dimensional figures presented were generated using Molscript (Figure 1a), PyMol (Figure 1b) and the topology diagram rendered in TopDraw (see supplementary information).

Circular Dichroism Spectroscopy. Protein samples were dialysed against 5 mM phosphate buffer overnight at 4 °C. Far-UV CD spectra and the corresponding blanks were recorded in a cuvette of path length 0.2 cm using a Jasco J-810 Spectropolarimeter by averaging 8 accumulations recorded at a rate of 10 nm/min, with a pitch of 0.5 nm, a bandwidth of 1 nm and a response time of 2 s. Near-UV spectra were recorded using a 1 cm cell, with a pitch 0.2 nm and a response time of 1 s. After subtraction of the appropriate blank, binominal smoothing was carried out within the Jasco Spectra Analysis program. Smoothed data were analyzed for protein secondary structure using the CDSSTR, SELCON3 and CONTIN/LL programs (41), accessed either *via* the Dichroweb service (42, 43) or the CDPro package. Both a general protein (SP43/dataset 4) and a membrane protein (SMP56/dataset 10) reference set were used (41, 44).

Full Methods are available in the online version of the paper at www.pnas.org.

Acknowledgements

The authors thank Professor E.N. Baker, University of Auckland for support and advice on the preparation of this manuscript, and the ESRF and SSRL for X-ray data collection facilities. V.A.M. thanks the European Molecular Biology Organization for a Long Term Fellowship. The work was supported in part by contributions from OneNorthEast and from the Royal Society. We

also wish to thank Dr G. Sharples and Dr E. Pohl, Durham University, for critical reading of the manuscript.

References

1. Fuentes, S., Tran, K. C., Luthra, P., Teng, M. N. & He, B. (2007) Function of the respiratory syncytial virus small hydrophobic protein *J Virol* **81,** 8361-6.
2. Collins, P. L., Chanock, R. M. & Murphy, B. R. (2001) in *Fields Virology*, eds. Knipe, D. M., Howley, P. M., Griffin, D. E., Lamb, R. A., Martin, M. A., Roizman, B. & Straus, S. E. (Lippincott-Raven, Philadelphia, Pa), Vol. 1, pp. 1443-1486.
3. Murphy, L. B., Loney, C., Murray, J., Bhella, D., Ashton, P. & Yeo, R. P. (2003) Investigations into the amino-terminal domain of the respiratory syncytial virus nucleocapsid protein reveal elements important for nucleocapsid formation and interaction with the phosphoprotein *Virology* **307,** 143-53.
4. Murray, J., Loney, C., Murphy, L. B., Graham, S. & Yeo, R. P. (2001) Characterization of Monoclonal Antibodies Raised against Recombinant Respiratory Syncytial Virus Nucleocapsid (N) Protein: Identification of a Region in the Carboxy Terminus of N Involved in the Interaction with P Protein *Virology* **289,** 252-61.
5. Henderson, G., Murray, J. & Yeo, R. P. (2002) Sorting of the respiratory syncytial virus matrix protein into detergent-resistant structures is dependent on cell-surface expression of the glycoproteins *Virology* **300,** 244-54.
6. Brown, G., Jeffree, C. E., McDonald, T., Mc, L. R. H. W., Aitken, J. D. & Sugrue, R. J. (2004) Analysis of the interaction between respiratory syncytial virus and lipid-rafts in Hep2 cells during infection *Virology* **327,** 175-85.
7. Takimoto, T. & Portner, A. (2004) Molecular mechanism of paramyxovirus budding *Virus Research* **106,** 133-145.
8. Peebles, M. E. (1991) in *The Paramyxoviruses* (Plenum Press, New York), pp. 427-456.

9. Chazal, N. & Gerlier, D. (2003) Virus entry, assembly, budding, and membrane rafts *Microbiol Mol Biol Rev* **67,** 226-37.

10. Ghildyal, R., Baulch-Brown, C., Mills, J. & Meanger, J. (2003) The matrix protein of Human respiratory syncytial virus localises to the nucleus of infected cells and inhibits transcription *Arch Virol* **148,** 1419-29.

11. Rodriguez, L., Cuesta, I., Asenjo, A. & Villanueva, N. (2004) Human respiratory syncytial virus matrix protein is an RNA-binding protein: binding properties, location and identity of the RNA contact residues *J Gen Virol* **85,** 709-719.

12. Ruigrok, R. W., Schoehn, G., Dessen, A., Forest, E., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. (2000) Structural characterization and membrane binding properties of the matrix protein VP40 of Ebola virus *J Mol Biol* **300,** 103-12.

13. Scianimanico, S., Schoehn, G., Timmins, J., Ruigrok, R. H., Klenk, H. D. & Weissenhorn, W. (2000) Membrane association induces a conformational change in the Ebola virus matrix protein *Embo J* **19,** 6732-41.

14. Timmins, J., Ruigrok, R. W. & Weissenhorn, W. (2004) Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogues but not homologues *FEMS Microbiol Lett* **233,** 179-86.

15. Han, Z., Boshra, H., Sunyer, J. O., Zwiers, S. H., Paragas, J. & Harty, R. N. (2003) Biochemical and functional characterization of the Ebola virus VP24 protein: implications for a role in virus assembly and budding *J Virol* **77,** 1793-800.

16. Jayakar, H. R., Jeetendra, E. & Whitt, M. A. (2004) Rhabdovirus assembly and budding *Virus Research* **106,** 117-132.

17. Schmitt, A. P., Leser, G. P., Waning, D. L. & Lamb, R. A. (2002) Requirements for budding of paramyxovirus simian virus 5 virus-like particles *J Virol* **76,** 3952-64.

18. Sugahara, F., Uchiyama, T., Watanabe, H., Shimazu, Y., Kuwayama, M., Fujii, Y., Kiyotani, K., Adachi, A., Kohno, N., Yoshida, T. & Sakaguchi, T. (2004) Paramyxovirus

Sendai virus-like particle formation by expression of multiple viral proteins and acceleration of its release by C protein *Virology* **325,** 1-10.

19. Gaudin, Y., Barge, A., Ebel, C. & Ruigrok, R. W. (1995) Aggregation of VSV M protein is reversible and mediated by nucleation sites: implications for viral assembly *Virology* **206,** 28-37.
20. Riffel, N., Harlos, K., Iourin, O., Rao, Z., Kingsman, A., Stuart, D. & Fry, E. (2002) Atomic Resolution Structure of Moloney Murine Leukemia Virus Matrix Protein and Its Relationship to Other Retroviral Matrix Proteins *Structure* **10,** 1627-1636.
21. Hatanaka, H., Iourin, O., Rao, Z., Fry, E., Kingsman, A. & Stuart, D. I. (2002) Structure of equine infectious anemia virus matrix protein *J Virol* **76,** 1876-83.
22. Gaudier, M., Gaudin, Y. & Knossow, M. (2002) Crystal structure of vesicular stomatitis virus matrix protein *EMBO J* **21,** 2886-92.
23. Dessen, A., Forest, E., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. (2000) Crystallization and preliminary X-ray analysis of the matrix protein from Ebola virus *Acta Crystallogr D Biol Crystallogr* **56,** 758-60.
24. Dessen, A., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. (2000) Crystal structure of the matrix protein VP40 from Ebola virus *EMBO J* **19,** 4228-36.
25. Holm, L. & Sander, C. (1996) Mapping the protein universe *Science* **273,** 595-603.
26. Gibrat, J., Made, T. & Bryant, S. (1996) Surprising similarities in structure comparison. *Current Opinion in Structural Biology* **6,** 377-385.
27. Krissinel, E. & Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions *Acta Crystallogr D Biol Crystallogr* **60,** 2256-68.
28. Faaberg, K. S. & Peeples, M. E. (1988) Association of soluble matrix protein of Newcastle disease virus with liposomes is independent of ionic conditions *Virology* **166,** 123-32.

29 Greenfield, N.J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 1, 2876-2890

30. Hoenen, T., Volchkov, V., Kolesnikova, L., Mittler, E., Timmins, J., Ottmann, M., Reynard, O., Becker, S. & Weissenhorn, W. (2005) VP40 Octamers Are Essential for Ebola Virus Replication *J. Virol.* **79,** 1898-1905.

31. Hansen, J. C., Lu, X., Ross, E. D. & Woody, R. W. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains *J Biol Chem* **281,** 1853-6.

32. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N. & Dunker, A. K. (2007) Intrinsic disorder and functional proteomics *Biophys J* **92,** 1439-56.

33. Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z. & Dunker, A. K. (2007) DisProt: the Database of Disordered Proteins *Nucleic Acids Res* **35,** D786-93.

34. Dombrowsky, H., Clark, G. T., Rau, G. A., Bernhard, W. & Postle, A. D. (2003) Molecular species compositions of lung and pancreas phospholipids in the cftr (tm1HGU/tm1HGU) cystic fibrosis mouse *Pediatr Res* **53,** 447-54.

35. Palestini, P., Calvi, C., Conforti, E., Botto, L., Fenoglio, C. & Miserocchi, G. (2002) Composition, biophysical properties, and morphometry of plasma membranes in pulmonary interstitial edema *Am J Physiol Lung Cell Mol Physiol* **282,** L1382-90.

36. Calculated using the ASC program; Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) J Comp Chem 16:273-284

37. Landolt-Marticorena, C., Williams, K. A., Deber, C. M. & Reithmeier, R. A. (1993) Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins *J Mol Biol* **229,** 602-8.

38. Sanderson, J. M. (2005) Peptide-lipid interactions: insights and perspectives *Org Biomol Chem* **3,** 201-12.

39. Ulmschneider, M. B. & Sansom, M. S. (2001) Amino acid distributions in integral membrane protein structures *Biochim Biophys Acta* **1512,** 1-14.

40. Sanderson, J. M. & Whelan, E. J. (2004) Characterisation of the interactions of aromatic amino acids with diacetyl phosphatidylcholiney *Physical Chemistry Chemical Physics* **6,** 1012-1017.

41. Sreerama, N. & Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set *Anal Biochem* **287,** 252-60.

42. Lobley, A., Whitmore, L. & Wallace, B. A. (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra *Bioinformatics* **18,** 211-2.

43. Whitmore, L. & Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data *Nucleic Acids Res* **32,** W668-73.

44. Sreerama, N. & Woody, R. W. (2004) On the analysis of membrane protein circular dichroism spectra *Protein Sci* **13,** 100-12.

Figure Legends

Figure 1: Three-dimensional structure of the Respiratory Syncytial Virus M protein. The crystal structure of M^{254R} (resolution 1.6 Å) shows two domains composed largely of β-sheets. Statistical information for the X-ray data is provided in the supplementary information. a) Divergent (wall eyed) stereo view of M^{254R} colored according to domain with the linker shown in cyan, the N-terminal domain in blue and the C-terminal domain in red. Residue R254 is shown in ball-and stick representation; b) A topology diagram of the protein. The linker between the N- and C-terminal domains is shown in magenta. Residues (numbers refer to Met as +1) in β-sheets are represented by broad arrows and helices as cylinders. c) Electrostatic surface potential (calculated with APBS) for M^{254R}, presented in a colour range from red to blue (-5 to +5 kT/e); uncharged residues are uncolored.

Figure 2: Comparison of the RSV M protein topology with that of Ebola VP40.
The cartoons show an overlay of the β-sheet arrangements of M^{254R} with EBOV VP40. A) shows RSV M protein N-terminal domain in blue and VP40 in yellow and B) shows M^{254R} protein C-terminal domain in red and VP40 in cyan. The same images are presented in supplementary information; Figure S2, in stereoscopic views.

Figure 3: CD spectrum of RSV M protein. Data sets of the far-UV (main box, 190 nm – 250 nm) and near- UV (inset box, 270 nm - 320 nm, axes are the same as the main plot) spectra were collected as indicated in the methods section. The data were analyzed for secondary structure information using the CDSSTR, SELCON3 and CONTIN/LL programs. A more comprehensive data set on *M254R* structure in solution is provided in the supplementary information (Supplementary figures S3, S4 and supplementary tables S2 and S3) and see also Table 1. A comparison with X-ray data with calculated CD spectra indicates that the crystal is more ordered than the solution structure.

Figure 4: Distribution of residues over the surface of RSV M that display significant sequence diversity from related proteins. a) Residues with a Q score ≤ 20 after alignment of all 8 pneumovirus proteins to M^{254R} are displayed in red (non-linker residues) or black (linker residues). The C-terminus is indicated by **C**; b) Residues with a Q score ≤ 20 or ≤ 50 after alignment of all RSV proteins are shown in red and blue respectively. The orientation of the structure is the same as part a). Alignments are presented in supplementary Figure S5

Figure 5: Distribution of tyrosine residues in the C-terminal domain of the RSV M protein. The N-terminal domain of M^{254R} has been omitted for clarity. The planar distribution of the residues (shown as stick and ball representations) on the surface of the M protein is readily apparent. The C-terminal residue is indicated by C; panels a) and b) show orthogonal views of this domain.

Table Legends

Table 1: Comparison of the solution and crystallographic structure of the the RSV M protein. (A) Summary of the output resulting from analysis of the M^{254R} structure using the Stride program. (B) Comparison of the CD and X–ray data. The output from the Contin/LL program is presented as this gave the best fit between experimental and computed structures. See supplementary information for more details of the various programs used to fit the CD data to known structures. The CD spectrum shows that the solution structure is more disordered that that of the crystallized form, compare “strand” (23 versus 47 %) and “coil” (31 versus 12).